Impaired Fertility and Perceived Difficulties Conceiving in Ghana: Measurement Problems and Prospects

Running Head: Measurement of Impaired Fertility in Ghana

Jasmine Fledderjohann*
Department of Sociology
University of Oxford
jasmine.fledderjohann@sociology.ox.ac.uk

David R. Johnson
Department of Sociology
The Pennsylvania State University

*Corresponding author

Abstract

Background: What is the most appropriate measure of impaired fertility for understanding its social consequences in sub-Saharan Africa (SSA)? While previous research has pointed to serious psychosocial consequences of infertility in SSA, the dearth of subjective measures of difficulties conceiving in surveys in the region has prevented comparisons of subjective and objective measures. Perceived difficulties conceiving may have a greater impact than objective measures for social outcomes such as divorce, stigmatization, and distress.

Methods: This paper assesses the most appropriate objective measure of impaired fertility for social research by comparing modified measures from the biomedical and demographic literature with a subjective measure of difficulties conceiving. Correlations are used to examine the relationship between measures, random effects models to explore how measures relate to background demographic characteristics, and test-retest models are employed to analyze the stability of measures across time. Secondary panel data collected in Ghana between 1998 and 2004 are used. Women aged 15 to 49 of all marital statuses were sampled (N=1,350). Correlations were examined for each measure with itself and with all other measures across waves. Random effects models were used to examine the relationship between measures and sociodemographic background characteristics.

Results: Longer waiting times to identification of impaired fertility required by demographic measures result in more stable measures; however, perceived difficulties conceiving are most closely aligned with clinical infertility (r=0.61; p<0.05). Comparing measures which have been adjusted for contraceptive use to unadjusted measures reveals that a large proportion of those identified as haivng impaired fertility based purely on waiting times are successful contraceptors. As a validity check, clinical infertility increases with age (OR=1.17; p<0.001) and other sociodemographics.

Conclusions: Where subjective measures are not available, epidemiological (24 month) measures may be most appropriate for studies of the social consequences of impaired fertility. Demographic measures are more stable; however, perceived difficulties conceiving, which are arguably most salient for social outcomes, are more closely related to epidemiological measures of impaired fertility. Accounting for contraceptive use is important in order to avoid false positives. Future research should consider a variety of measures of perceived difficulties conceiving and self-identified infertility to assess which is most valid; in order to accomplish this, it is imperative that subjective measures of infertility be included in social surveys in SSA.

Word Count:

Introduction

Involuntary childlessness and difficulties conceiving are associated with detrimental psychological and social effects in a variety of contexts across sub-Saharan Africa (SSA) (Dyer, Abrahams, Hoffman, & Van der Spuy, 2002; Dyer, Abrahams, Mokoena, Lombard, & van der Spuy, 2005; Fledderjohann, 2012; Hollos & Larsen, 2008; Hollos, Larsen, Obono, & Whitehouse, 2009; Rouchou, 2013). In light of the negative social consequences associated with infertility, the high rates (Larsen, 2000; Mayaud, 2001) found across much of the subcontinent are troubling. However, without adequate measures of infertility, the social impact of the condition is difficult to pinpoint. Much of the literature on measurement of infertility in SSA focuses on the best measure for estimating prevalence, and on the utility of demographic measures in particular (Larsen, 2000, 2005; Larsen & Menken, 1989, 1991; Larsen & Raggers, 2001), but these measures very often err on the side of producing conservative estimates, and may not align with women's own defintions of infertility (Gerrits, 1997; Leonard, 2002). More broadly, as several previous studies have noted (Gnoth et al., 2005; Gurunath, Pandian, Anderson, & Bhattacharya, 2011; Habbema et al., 2004), there is a wide variety of measures available in the extant literature. A key criticism of extant measures is that even measures falling within the same broad category (e.g. epidemiological measures, demographic measures) vary from one another on factors such as how long the couple must spend trying to conceive, whether individuals should be excluded from definitions of infertility on the basis of their marital status, and age for inclusion in the sample. One consequence of this ambiguity is substantial disagreement on point estimates of infertility prevalence (Gurunath et al., 2011).

Following from this criticism, a broad debate has emerged regarding how infertility should be measured, particularly in the context of comparability when assessing prevalence (Gnoth et al., 2005; Gurunath et al., 2011; Habbema et al., 2004). We would add to this debate that comparability is of value

beyond the context of prevalence studies, and that assessments of the social consequences of infertility are improved by consistency with prevalence figures, and vice versa: if the social consequences of infertility are found to be severe, it is helpful to know what fraction of the population is impacted. Inversely, if there are few consequences of infertility, a high prevalence may not be of great public health concern when compared to other reproductive health issues. In the context of limited public health resources, consistency of measurement is thus crucial. Furthermore, as some of the highest rates of infertility in the world are found in SSA (Mascarenhas et al., 2013), we contend that a suitable measure should be capable of being constructed with relative ease from secondary survey data, as much of the quantitative work on infertility across SSA (and in other low and middle income regions, where infertility rates are also high) is based on secondary survey data. Finally, we argue that, in addition to underlying biological ability to conceive, perceptions about one's ability to conceive are likely to be salient for understanding social outcomes (Leonard, 2002), and a useful measure of infertility should therefore also correlate with subjective assessments. Using Ghana as a case study, in this paper we compare several potential measures of infertility with the aim of identifying a measure which a) closely matches standard objective measures, b) can be constructed from survey data, and c) is correlated with subjective assessments of ability to conceive.

Childbearing in Ghana

Childbearing, which provides couples with economic resources, adult status, and ancestral ties, is an expected and nearly universal aspect of marriage in Ghana; consequences of infertility are often severe (Donkor, 2008; Farnes, Beckstrand, & Callister, 2011; Oppong & Abu, 1987; Tabong & Adongo, 2013). The practice of bridewealth is on the decline in Ghana, but, where still practiced, repayment may be demanded of the bride's family if the couple fails to produce biological offspring (Armstrong, 1997;

Aryee, 1997), highlighting the centrality of childbearing to marriage in Ghana (and placing severe pressure to conceive on the bride). In addition to suffering repayment of the bridewealth, infertile women may also lose access to their land and homes, and may struggle to find a source of economic support (Hollos et al., 2009). Beyond economic concerns, infertile women are often blamed for their infertility, resulting in substantial social stigma and ostracism (Fledderjohann, 2012; Tabong & Adongo, 2013). Infertile women may be labelled as worthless, and are seen as incapable of fulfilling their roles as adult women, giving rise to depression and diminished self-worth (Donkor, 2008). Childlessness may even lead to denial of important funerary and burial rights in some contexts, particularly where funeral costs are borne by one's offspring (Donkor, 2008; Hollos et al., 2009).

Although the majority of childbearing occurs within the context of marriage, non-marital childbearing is not uncommon (Gyimah, 2003; Takyi & Gyimah, 2007). Of particular interest in the context of this study, some extramarital partnerships may in fact be entered into for the specific purpose of testing one's fertility if a child has not been conceived within the marriage (Fledderjohann, 2012). While family size is declining, the Ghanaian total fertility rate (TFR) of 4.2 children per woman is low compared to the figure of 5.1 children for SSA as a whole (Population Reference Bureau, 2012). The timing and pace of fertility are influenced by factors such as norms regarding postpartum abstinence and breastfeeding, gender roles, and migration patterns. These factors in turn shape the duration of spells of abstinence, coital frequency, fertility desires, contraceptive use, and other determinants of fertility (Bongaarts, Frank, & Lesthaeghe, 1984).

However, the associations between fertility and its determinants are not straightforward, and often these factors interact in complex ways. For example, where there is not a high premium placed on gender egalitarianism in relationships, partner fertility preferences may asymmetrically influence fertility behaviors. Several studies of contraceptive use and fertility behavior in Ghana have provided evidence

that a sizeable portion of the inverse relationship between women's empowerment and fertility rates in fact operates through the influence of partner preferences in fertility behavior (DeRose & Ezeh, 2005; Dodoo & Frost, 2008; Ezeh, 1993). Similarly, pointing to the complexities of the determinants of fertility in Ghana, a recent study of urbanization and fertility provides evidence of a strong, negative association between urbanization and fertility (White et al., 2008), but also shows that this association differs by parity, and may in part be driven to selection effects, in which those most likely to migrate to urban areas may already have lower fertility desires. These examples highlight the multifaceted nature of family structure and childbearing decisions in Ghana, and point to the need for an infertility measure that is applicable across a wide variety of family formation pathways.

Measurement of Infertility

Measurement of infertility is an area of much debate in recent literature, owing in part to the vast and growing array of available measures (Gnoth et al., 2005; Gurunath et al., 2011; Habbema et al., 2004). While there is wide agreement that a standarized, comparable measure of infertility is needed, there is less agreement about the form the measure should take. Habbema et al. (2004) argue that ambiguous terminology in the infertility literature has led to confusion of infertility with sterility, as well as disagreement about how long a couple must try to conceive before being labelled infertile, leading to confusion both within the literature and, in a medical setting, among patients. To ameliorate these issues, they advocate using a three-pronged clinical approach based on a descriptive statement of time spent trying to conceive, a diagnostic measure of ability to acheive a pregnancy (rather than live birth), and a prognostic statement on a 4 point scale ranging from normal fertility (Grade 0) to sterility (Grade 4) based on the probability of spontaneous conception. Gnoth et al. (2005) agree that a clinical measure is useful, but instead propose a 3-step grading of infertility: slightly subfertile (after 6 ovulatory cycles without a

conception), moderately or serious subfertile (after 12 cycles—the standard cut-off in the literature), and completely infertile (after 48 cycles). Notably, while the description of both approaches includes a discussion of the decline in ability to conceive with age, neither advocates the exclusion of couples or women on the basis of a strict age cut-off. In fact, Gnoth et al. (2005) assert that the effect of age varies on the basis of individual factors to such an extent that the association between probability of conception and age is only able to be assessed after a first clinical infertility exploration.

The approaches outlined above are clearly aimed at medical practitioners, and may be of less use to researchers focussed on the prevalence and/or psychosocial implications of infertility rather than on its treatment. While we agree that consistency of measurement is vital for ensuring the comparability of findings, we contend much of the research on infertility in low and middle income countries comes from secondary survey data, and clinical measures may be difficult to construct from these data. Especially concerning is that survey responses cannot adequately capture conception (Larsen, 1994), and may underestimate pregnancy loss (J. B. Casterline, 1989). In order to facilitate cross-national comparisons of the prevalence of infertility over time, Gurunath et al. (2011) call for development of a standarized definition of infertility based on an age-specific waiting time to pregnancy. They note that, while biomedical measures of infertility fail to capture the spectrum of reproductive capacity (conception may be more or less difficult, rather than simply a binary state) and are of limited utility for prevalence and public health research, demographic measures are not useful for identifying patients in need of treatment in a clinical context owing to the long waiting times to conception required by the measures.

In the biomedical literature, clinical definitions identify infertility as no conception after 12 months of regular, unprotected sex; epidemiological definitions extend the requisite length of intercourse to 24 months (Marchbanks, Peterson, Rubin, & Wingo, 1989), but the focus remains on conception rather than live births. Clinical definitions are often used in medical settings, where short waiting times to

diagnosis and a focus on conception ensure that couples will receive treatment early to address potential problems. These measures may overestimate infertility, as couples may naturally take longer than twelve months to conceive without being infertile (Larsen, 2005). Demographic infertility measures are based on live births rather than conception for a sexually active woman not using contraception (Larsen, 2005), and most require 5 or 7 years from either (a) the date of marriage in the case of primary infertility (childlessness), or (b) the date of last birth in the case of secondary infertility (subsequent to the birth of a child). While this approach could still overestimate infertility rates, the long waiting times used for these measures tend to minimize bias.

When contraceptives are used, lack of pregnancy likely reflects the effectiveness of the contraceptive method rather than underlying infertility. While some couples using contraceptives may be sterile, failure to remove successful contraceptors from the pool of women at risk for infertility results in an overestimate of infertility (Larsen, 1994; Marchbanks et al., 1989). Modern contraceptives may be more effective than traditional methods at preventing pregnancy, but research in Tanzania and Nigeria suggests that some women may believe that modern contraceptives actually cause infertility (Gijsels, Mgalla, & Wambura, 2001; Koster-Oyekan, 1999; Mgalla & Boerma, 2001). Women who desire no more children may also be consciously limiting fertility via methods other than contraceptive use (such as lactational ammenhorea), but women's fertility desires may have a limited impact on their fertility behavior due, in part, to the strong role of men's desires in shaping fertility behavior (DeRose & Ezeh, 2005; Ezeh, 1993). Women who are infertile with intent (i.e. who desire to have children but are infertile) report greater distress and other negative consequences than women without intent (that is, who are sterile but not trying to conceive; (Greil & McQuillan, 2004; Greil, McQuillan, Johnson, Slauson-Blevins, & Shreffler, 2010). Although excluding those who do not wish to conceive from the pool of infertile women may underestimate infertility, including those who are infertile without intent may cause a downward bias

in estimates of the social impact of infertility. There is a natural decline in the ability to conceive over time (Broekmans, Knauff, te Velde, Macklon, & Fauser, 2007; Gnoth et al., 2005; Gurunath et al., 2011; Larsen, 1994), and infertility can be expected to increase as cohorts age. Whether perceived difficulties conceiving also increase as cohorts age is unclear. There may also be religious and ethnic differences in the prevalence of infertility, perhaps as a result of differences in sexual and childbearing norms (Anarfi & Owusu, 2011; Ericksen & Brunette, 1996; Frank, 1983; R. G. White, Zaba, Boerma, & Blacker, 2001).

The Value of Self-Identified Infertility

Little attention has been paid to the association of biomedical and demographic measures with perceived ability to conceive, particularly in SSA, due in part to the dearth of self-identified measures in survey data. However, there is some evidence pointing to the value of self-identified measures in the study of infertility. One study (Dick et al., 2003) using case-control data on ovarian cancer in Australian women showed only 70% of women who were identified as infertile using an epidemiological measure self-identified as infertile, and 5% who were identified as fertile reported difficulty conceiving. The study concluded that self-reported difficulty conceiving is a particularly useful measure for assessing the burden of fertility problems. In Tanzania, Larsen (2005) found that a significantly higher proportion of women were identfied as infertile based on time to conception than on time spent trying to conceive (in both cases using a 24 month measure), providing further evidence that perceptions may not align with biomedical definitions as currently constructed.

Yet it remains unclear how perceived difficulties conceiving and more objective measures in the biomedical and demographic literatures are related. For many outcomes of interest, perceived difficulties conceiving may be more salient than externally defined measures. It may be the case that self-assessment

is the most useful measure for understanding the influence of infertility on social outcomes; the effects will presumably be greatest when a woman is *perceived* to be infertile, by herself and/or her husband, family, friends, and other members of her community. For example, marital discord will likely arise only when difficulties conceiving are acknowledged by at least by one member of the dyad. Indeed, US-based survey research has shown that, compared to other measures of fertility barriers, self-identified infertility is the greatest predictor of fertility-specific distress (Jacob, McQuillan, & Greil, 2007). Acceptable waiting times to pregnancy—that is, the amount of time a couple may try to conceive before suspecting impaired fertility—will differ between social groups (Leonard, 2002), reducing the stability of the measure and thereby making the subjective assessments less attractive for assessing the epidemiological prevalence of infertility in a population. Yet "objective" measures in biomedicine and social science may fail to accurately capture the social consequences of infertility, as these measures apply a uniform waiting time to all women, and are therefore unable to account for variability in personal and cultural perceptions as to what constitutes an acceptable waiting time to conception.

Leridon's (1991) work on impatience to conceive in France during 1978-1988 highlights the importance of perceptions in shaping who is identified as infertile. Leridon documents a steep rise in self-reported difficulties conceiving, without corresponding evidence of a rise in infertility, which he attributes to changing norms around acceptable waiting times. In SSA, studies which focus on perceptions of infertility are rare due to data limitations; however, a few key qualitative studies provide some evidence on the importance of these perceptions. Using in-depth interviews with 26 Malawian men and women, Barden o'Fallon (2005) finds that expectations for rapid childbearing are high, and that couples who do not conceive within a period as short as a few months may begin to preceive fertility problems. Based on qualitative interviews in Chad, Leonard (2002) notes that the underlying assumption of demographic measures that infertility is the same across time and place is inaccurate, and argues instead that infertility

may be perceived as having too few children, with the definition of "too few" varying across time and between cultures. As a result, Leonard contends, subjective measures of infertility may be a better predictor of social outcomes, including marital disruption, abuse, poverty, social isolation, and treatment-seeking.

Contributions of the Current Study

Work on perceived difficulties conceiving in SSA suggests that self-identified measures may be more salient for social outcomes, but this body of work tends to rely on small samples of cross-sectional data, and has generally not examined the degree to which subjective measures are correlated with more objective measures from the demographic and biomedical literature. Acceptable waiting times to pregnancy, which are shaped by cultural norms (Gijsels et al., 2001; Leonard, 2002), influence who identifies as having difficulties conceiving. An accurate understanding of the implications of infertility in SSA requires use of an appropriate definition: if objective and subjective assessments are closely aligned, it may be possible to use the measures interchangeably in social science research; if, however, the correlation between subjective and objective assessments is low, reliance on biomedical/demographic measures may provide an inaccurate picture of the social consequences of infertility in the region. In order to assess the utility of these measures, this paper uses longitudinal data to examine: 1) which measure of infertility that can be constructed from survey data is tied to both objective biomedical/demographic measures and perceived difficulties conceiving and, as a validity check, 2) whether measures relate to background characteristics (age, ethnicity, religion) in an expected manner. We base our assessment of relationship to background characteristics on the association between sexual practices, infertility, and sociodemographic background established in previous literature.

Methods

Data and Measures

Eight rounds of secondary longitudinal data collected across 6 communities in Ghana by the Population Council of New York and the University of Cape Coast between 1998 and 2004 were used for the analyses. Ethics approval for this research was obtained from the Pennsylvainia State University Institutional Review Board. Selection of the six communitities included in the sample was based on a purposive sampling design in which between-community diversity in local ecology, economic modality, and ethnicity was maximized (for full details, see J. Casterline, 2007). Preliminary data collection to scope the proposed survey sites in the form of exploratory interviews and focus groups, a pilot survey, and a household census took place between 1994 and 1997. Of the 6 communities, 2 were in coastal and 4 were in inland regions. The locations were mainly rural, located in the Western, Central, and Greater Accra regions, and the primary economic modalities across communities included a range of fishing, trading, and farming. Within the 4 smallest communities, the households were enumerated in the preliminary data collection phase, and all women of reproductive age were included in the survey sample. In the larger communities, simple random sampling was employed to select the sample. As a result of this survey design, the data cannot be taken to be representative of Ghana as a whole, but are representative of the selected communities. Interviews were conducted face-to-face in local languages (Ewe, Fante, Ahanta, Twi, Ga, Adangbe, or Hausa) using an instrument that had been pretested both in English and in the local language. Interview teams consisted of 1 supervisor and 4 interviewers, with interviewers being selected from the sampled communities.

Women aged 15 to 50 at wave 1, of all marital statuses, were sampled. Respondents were given a main survey relating to demographic characteristics, fertility behaviors, and other variables. Respondents also provided retrospective information on contraceptive use for each month between waves. Due to left censoring, data were not available for contraceptive use prior to the first wave of data collection. In wave 1, 1,219 women were sampled; 219 women were added in wave 2 to adjust for attrition. Twelve cases attrited, resulting in a sample of 1,364. Where respondents did not participate in one wave, considerable effort was made to locate the respondents in the next wave, resulting in some small fluctuation in sample sizes from wave to wave. The original Cape Coast survey schedule and sample sizes at each wave are provided in Supplemental Figure 1. In order to compare the same sample of women across measures while reconciling quite a disparate array of infertility criteria, we restricted the sample to women aged 15-49 because the association between self-identified difficulties conceiving and age is not well-established in the literature, and because many survey data sets focussed on reproductive health tend to focus on this age range. The final analytic sample size was 1,350. Missing data for background and demographic variables was around 3% in most cases, and more varied among other indicators. Fertility desires had the highest amount of missing data (19.27%). Missing data were multiply imputed using the ICE procedure in Stata 11 with 10 imputed datasets. Results were combined using the *mim* procedure.

We draw on four objective measures of infertility from the literature, but with important modifications to fit with the goals of identifying a measure that a) closely matches standard objective measures, b) can be constructed from survey data, and c) is correlated with subjective assessments of ability to conceive. Note, we use the term "standard" measure loosely here given that, as discussed above, there is considerable disagreement in the literature on how to construct these measures. The main criterion we used to maintain consistency with previous literature was time to a live birth. Most common in the extant literature is a 12 month measure, typically called clinical infertility, followed by 24, 60, and 72

month measures (Gurunath et al., 2011). We did not make any adjustments for postpartum amenorrhea, both because specific questions on amenorrhea may not be available in survey data, and because it may be difficult to distinguish postpartum from nutritional amenorrhea, which may be of particular concern in countries with high rates of malnutrition.

The objective measures modified and included here were clinical, epidemiological, demographic 5 year, and demographic 7 year infertility. These measures of infertility focused on women who were infertile with intent (Greil & McQuillan, 2004; Greil et al., 2010). Unfortunately, it was not possible to distinguish between primary and secondary infertility due to data limitations. Specifically, the number of cases of primary infertility was too low (as low as 7 cases by the end of the survey in the case of the demographic 7 year measure) to support models split by primary and secondary infertility. The focus was on live births due to difficulties accurately identifying conception (particularly pregnancy wastage, which may be underestimated by 50% in survey data; J. B. Casterline, 1989). Additionally, survey data may underestimate pregnancy if the respondent is pregnant but not aware of the pregnancy. This is particularly problematic for cross-sectional data, where there is no opportunity for a follow-up survey to reveal a pregnancy retrospectively. Moreover, measurement of conception is vital in a clinical setting, where the etiology of infertility determines the available treatment. However, as infertility is a social phenomenon in addition to being a biological condition, a focus on live births would seem more appropriate, as the goal for most couples is a live birth, not simply a conception. In fact, in some settings, even having too few children, or not having children of a particular sex may be viewed as a form of infertility (Leonard, 2002).

The clinical definition identified a woman as infertile after 12 months of regular, unprotected sex without a birth; the epidemiological definition extended the waiting time to 24 months. The demographic definitions identified a woman as infertile if she had not achieved a live birth after 5 or 7 years of unprotected intercourse. The distinguishing factor between objective measures was waiting time. Self-

assessed difficulties conceiving were measured by responses to the question "When you want to become pregnant, do you become pregnant quickly, or does it take a long time? " Women who responded "Takes a long time" or "Can no longer become pregnant" were classified as having perceived difficulties conceiving; women who responded "Quickly" were classified as not reporting any difficulties conceiving. Women (5.4%) who responded "Cannot get pregnant" to a second question, "Would you like to have (a/another) child (with your husband/partner) or would you prefer not to have any (more) children (with him)? " were also classified as reporting perceived difficulties conceiving. These questions were asked only in waves 1 and 6-8.

Among respondents who perceived difficulties, 83.5% responded "Takes a long time" and 16.5% said that it is impossible (based on either measure) to conceive. The response "takes a long time" could mean that a woman takes longer than average to conceive but is capable of conception, is infertile but reluctant to classify herself as such, or is unaware that she is infertile. However, Greil (1991) found that U.S. women undergoing treatment were more likely to identify as "not yet pregnant" than "infertile," suggesting that indirect questions may include women who experience stigma and fertility-specific distress, but are relectant to embrace the infertile label. Similarly, qualitative work by Loftus (2009) and Andrews et al. (1991) shows that infertile women may experience the social consequences of infertility without labeling themselves as such due to the negative connotations associated with the label, supporting use of the response "takes a long time" as a measure of perceived difficulties conceiving.

A substantial portion (25.5%) responded "don't know" when asked whether they become pregnant quickly. While women who express uncertainty about their infertility status could conceivably say they don't know because they are not currently sexually active, they are unlikely to be identified as infertile because the contraceptive measure used includes abstinence. Moreover, only women who were married or in a union were included; while couples may abstain due to migration, norms about postpartum

intercourse, or a variety of other reasons, excluding single women from the sample does reduce the risk of identifying sexually inactive women as infertile. Robustness checks were conducted using correlations, test-retest, and random effects models to compare the effects of dropping respondents who replied "don't know," coding "don't know" responses as reported perceived difficulties conceiving, and coding "don't know" responses as reporting no difficulties. Based on these robustness checks, women who replied "don't know" were counted as reporting perceived difficulties. Waiting times were calculated based on months since the most recent birth. For women who had never given birth, waiting times were based on the union start date. Due to small cell counts for primary infertility (as discussed above), primary and secondary infertility were combined, with infertile women were coded 1, other as 0.

In order to avoid false positives in the coding of infertility, successful contraceptors should not be counted as infertile; however, it seems that no prior research has systematically compared different methods for accounting for contraceptive use—particularly when assessing the correlation between subjective and objective measures. In this study, three subsets of measures were created: (a) unadjusted measures, which do not account for contraceptive use or fertility desires, (b) contraceptive adjusted measures, which consider only women who desire to have a child and who are not using any contraceptives (including herbs, withdrawl, and abstinence) as potentially infertile (with women who do not desire additional children and/or who are currently contracepting coded as 0), and (c) modern adjusted measures, which treat only women who desire children and are not currently usingmodern contraceptive methods (IUD's, the pill, etc.) as potentially infertile. Given that equivalent results were obtained for contraceptive and modern adjusted measures, only results for modern adjusted measures (hereafter referred to as adjusted measures) are presented in the random effects and test-retest models. The risk period began at the first survey in which no contraceptive use was reported. Questions regarding current contraceptive use for each month were taken from the retrospective calendar data. Infertile women may

use contraceptives; while false negatives would create a downward bias in estimates of infertility, if infertility is unperceived, it is unlikely to have social consequences. Supplemental Table 1 provides an overview of how each measure was constructed, with differences in waiting times and contraceptive use between measures catalogued.

Contraceptive use was higher than expected. Among those who said it takes a long time to conceive, 65% reported using any form of contraceptive. Contraceptive use was even higher among those who say it is impossible to conceive (73%), but lower among those who said they don't know (56%). Contraception was highest among those who say it is impossible to conceive. This may stem from cultural notions about aging and reproductive fatigue (Bledsoe, 2002), and may support the notion that couples use contraceptives for purposes other than pregnancy prevention (Meekers & Calvès, 1997).

Analytic Strategy

First, correlations between measures of infertility/perceived difficulties and contraceptive use were examined at each wave and across waves. The next set of analyses examined relationships between measures of infertility, perceived difficulties, and key background variables: age, ethnicity, and religious identification. Age was expected to be positively associated with all measures of infertility, as prior research has documented the decline in the ability to conceive with age (Weinstein, Wood, & Chang, 1993). Ethnicity and religious affiliation were considered due to their association with sexual and birthing practices (Addai & Trovato, 1999), which may shape both exposure to risk factors for infertility (reproductive tract infections, STI's, etc.; Addai, 1999, 2000; Frank, 1983) and notions of acceptable waiting times. For this analysis, observations were pooled across all waves. As pooling violated the independence assumumption (Johnson, 1995), random effects logistic regression models were employed

in order to adjust for the non-independence of observations across waves (Hall, 2000; Hartzel, Agresti, & Caffo, 2001).

The paper also sought to answer the question of which measure was the most reliable over time using a test-retest model. A tetrachoric test-retest model for dichotomous outcomes was employed to examine the stability of measures across waves; as with the random effects models, the test-retest approach appropriately accounts for the non-independence of observations across waves (Alwin, 2007; Johnson, 1995). Assuming the least stringent biomedical measure, most stringent demographic-style measure, and the subjective assessment of difficulties conceiving will approximately capture the range of impaired fertility, and given the high correlation of biomedical and demographic measures, only one biomedical measure (clinical) and one demographic (7 year) measure are included in the model. Because the perceived difficulties measure was not available in waves 2 through 5, the test-retest analysis was limited to waves 6, 7, and 8. Wave 1 was excluded both because trends could not be continuously assessed throughout waves 2 through 5, and because impaired fertility figures in wave 1 are slightly inflated.

One advantage of using random effects models with longitudinal data as we have done here is that within-person change can be modelled. In the cross-sectional case, it is difficult to disentangle a cohort effect (for example, if improvements in reproductive healthcare across time have meant better reproductive health for younger individuals) from an age effect, in which the underlying biological ability to conceive diminishes with age. As the longitudinal models employed here track within-person change across time, the age covariate more accurately reflects the underlying phenomenon of interest. In a similar vein, the aim of the test-retest models is to uncover which measure is most reliable over time. A measure which accurately captures shifts in underlying ability to conceive is crucial, and such an assessment is best accomplished with longitudinal data.

Results

Supplemental Table 2 provides descriptive statistics. Respondents are aged between 15 and 49, with a mean of 32.33. Mean age at first marriage (19.47) is on par with national statistics (Demographic and Health Surveys, 2011). A majority (51%) of the sample is Fante. Over 60% of the sample is married, and an additional 15% are involved in a non-marital union. Twelve percent are never-married, while the remaining 10% are either separated, divorced, or widowed. Note, the inclusion of separate, divorced, and widowed women in the sample represents women who were married or in a union in previous waves, but who were excluded from subsequent waves after relationship disruption because they no longer fit the inclusion criteria. Nearly a quarter of the sample identifies as Moslem (22%), and an equal percent identify with a syncretic, traditional, or another religion. Only 8% report no religious affiliation.

Fifty-seven percent report wanting a(nother) child. Sixty-five percent of respondents use contraceptives, and 57% report using a modern method. Respondents average 3.53 children, close to the TFR for Ghana. Thirty-six percent of the sample reported that they become pregnancy quickly. Nearly another third (32%), say it takes a long time for them to become pregnant, while 6% say it is impossible. An additional 26% responded "don't know" to this question.

[Table 1 Here]

Table 1 presents the proportion infertile across waves for the measures included in subsequent analyses (for findings for measures not included in the subsequent analyses, see Supplemental Table 3). Impaired fertility figures in wave 1 are slightly inflated when compared to subsequent waves, likely due to an ambiguous measurement period for contraceptive use at wave 1. Similarly, the increase in impaired fertility observable in wave 8 may reflect women who are pregnant but do not realize it. Unadjusted and adjusted versions are presented for clinical, epidemiological, and demographic-style measures, as well as

for perceived difficulties conceiving. More women are identified as infertile by clinical measures than by epidemiological and demographic measures. The proportion identified as infertile by the unadjusted measures are substantially higher than adjusted measures.

With minor exceptions, impaired fertility increases across waves, likely as a result of aging across waves. The highest estimate, unadjusted clinical in wave 6, identifies nearly three quarters of the sample (.74) as infertile. The lowest estimate, the demographic-style 7 year measure (adjusted, in multiple waves), identifies only 1% of the sample as infertile. A large portion of this variation can be accounted for by eliminating successful contraceptors from the pool of at-risk women. Consider only the measures which exclude women who are using contraceptives: the highest proportion identified as infertile drops from .74 to .20 (adjusted perceived difficulties, waves 1 and 8). The second factor accounting for this variation is waiting time. Short waiting times may reflect *difficulty* conceiving without indicating *inability* to conceive; couples may still conceive naturally beyond the clinical 1 year mark (Larsen, 2000, 2005). For perceived difficulties conceiving, culturally defined "acceptable" waiting times may influence identification as infertile; variance between subjective and objective measures is partially dependent on cultural norms.

Correlations within and between all measures across all waves were tested. For brevity, selected results are presented here. Results for epidemiological infertility are remarkably similar to clinical results, while demographic-style 5 and 7 year results closely approximate one another. Among the unadjusted objective measures, impaired fertility and contraceptive use are positively, significantly associated within each wave; a significant portion of women who are classified as infertile by the unadjusted measure are successful contraceptors. There is a weak, sometimes significant, positive correlation between perceived difficulties conceiving and contraceptive use. Correlations were examined for each measure with itself across waves. For all measures, correlations range from weak to moderate, with correlations between

proximate waves being greater than those between distal waves. Associations between impaired fertility in wave 1 and subsequent waves are the weakest, likely due to the inflated frequencies discussed above. The same relationship holds for perceived difficulties conceiving. Correlations are generally positive and statistically significant, with strength ranging from weak to moderate.

The highest correlation for the biomedical measures across waves is .47. There appears to be substantial variation in who is identified as infertile from wave to wave, likely as a result of the short waiting times involved in these measures. Among demographic-style measures, the highest correlation is .60, with less variation from wave to wave, likely due to longer waiting times resulting in greater stability. Correlations between perceived difficulties are comparatively low; the highest correlation among perceived difficulties is .30. Among objective measures, correlations are highest among measures similar in terms of waiting times; the correlation between adjusted clinical and epidemiological measures is .86– nearly as high as the correlation between the clinical measure and its adjusted contraceptive counterpart. The correlation between the clinical and demographic-style 5 and 7 year measures are much lower, at .52 and .38 respectively. Perceived difficulties are most closely aligned with the clinical measure, followed by epidemiological, demographic 5 year, and, finally, demographic 7 year measures; perceived difficulties are positively associated with the objective measures, though the demographic 7 year measure is non-significant.

[Table 2 About Here]

Table 2 provides results for random effects models examining the relationship between measures and background characteristics. Model 1 shows clinical infertility as a function of age, ethnicity, and religious identification. For every one year increase in age, there is 9% increase in the odds of identification as infertile. Denkyira respondents have greater odds (OR=1.45) of identification as infertile

than the reference group (Ahanta or other ethnicity). Religion is not a significant predictor. Model 2 shows the results for adjusted clinical infertility. The relationship between adjusted clinical and age is highly significant, positive (OR=1.17), and curvilinear; ethnicity and religious affiliation are non-significant predictors.

Model 3 shows unadjusted demographic infertility as a function of age, ethnicity, and religious affiliation. Age is positively, significantly associated with infertility: for every year increase in age, the odds of identification as infertile increase by 18%. Denkyira respondents are significantly more likely (OR=1.75) to be infertile than those identifying as Ahanta or another ethnic group. The odds of infertility are 52% higher for Protestant respondents than those who don't identify with any religion. The results for the adjusted demographic measure are provided in Model 4. The relationship between age and standard demographic infertility is significant (OR=1.40) and curvilinear, with infertility increasing until around age 40, then declining slightly thereafter. Ethnicity and religion are no longer significant predictors.

The final models in Table 2 provide the random effects models for perceived difficulties conceiving. There is a statistically significant curvilinear relationship between unadjusted perceived difficulties conceiving and age. Model 5 shows that perceived difficulty conceiving declines until around age 30, then begins to rise steadily thereafter. Catholic respondents have 36% lower odds of perceiving difficulties conceiving than those with no religious affiliation. Model 6 shows the results for the adjusted measure. The relationship between age and perceived difficulties is significant but no longer curvilinear once contraception is accounted for. Additionally, accounting for contraception reduces the effect of religious affiliation to non-significance.

[Table 3 About Here]

Table 3 provides results for the test-retest model. Model fit, measured by the Root Mean Square Error of Approximation (RMSEA), was examined for the following theoretically motivated models: a) All errors were uncorrelated, and no paths were constrained to be equal across waves, b) errors for each measure between waves 6 and 7, and waves 7 and 8 are correlated, c) errors for each measure between waves 6 and 7, 6 and 8, and 7 and 8 are all correlated, d) errors were uncorrelated, but paths for each measure were constrained to be equal across waves, e) paths were constrained to be equal and errors were correlated between waves 6 and 7 and waves 7 and 8, and, finally, f) within measures across waves, all errors were correlated, and all paths were constrained to be equal. Based on the model fit statistics, model f was selected (RMSEA=.027). Results presented are from standardized models.

Coefficients within measures across waves have been constrained to be equal, and thus do not vary across waves. The coefficients for the impaired fertility measures represent their reliability (Alwin, 2007; Johnson, 1995). Clinical infertility appears to be the most reliable measure of the latent construct, as indicated by the large coefficients for this measure across waves, though the coefficient is not significant. Clinical infertility is followed by the demographic 7 year measure, which is a highly significant indicator of underlying infertility. Reliability is lowest for perceived difficulties conceiving, though also significant and only slightly less reliable than the demographic measure.

The paths for the relationship between latent impaired fertility in waves 6 and 7 and waves 7 and 8 represent the stability of the measures across waves. Although these paths were not constrained to be equal, they are remarkably similar, suggesting that the stability between waves 6 and 7 is only marginally lower than the stability between waves 7 and 8. This suggests that the stability of the measures fluctuates very little across time. However, given that these coefficients are only moderate in strength (B=.57 and .58 respectively), the stability of the measures is not particularly high across waves. This is somewhat surprising given within-person variation in biological ability to conceive could reasonably be expected to

be fairly low, and further pointing to the salience of psychosocial factors in identifying impaired fertility. The moderate stability of the measures across waves, then, suggests that current measures are imperfect indicators of underlying inability to conceive.

Finally, Table 3 includes the correlations of errors to capture associations among unmeasured characteristics. While these correlations do not provide information about the reliability or stability of the measures, they do suggest that there is a significant association between the unmeasured characteristics of respondents who are identified as infertile by the demographic measure and, to some extent, by the perceived difficulties conceiving, but not those who are clinically infertile. Characteristics used to identify demographic infertility vary little as a result of the long waiting time; this is less true of perceived difficulties conceiving, which does not specify a waiting time, and especially of clinical infertility, which designates a very short waiting time, causing substantial change in identification according to the latter measures across time. Moreover, because the time span of the demographic measures is greater than the time between waves, some of the same births are counted in each measure producing substantial autocorrelated error. Overall, the test-restest models show that the demographic 7 year measure was the most reliable statistically significant indicator of infertility. This provides some support for Larsen's (2005) assertion that biomedical definitions of impaired fertility are not stringent enough to be useful for estimating prevalence.

Robustness Checks

We tested a number of alternative specifications of our models as a robustness check on our findings. First, secondary infertility may differ qualitatively from involuntary childlessness, both in terms of the social consequences of infertility and the likelihood of self-identifying as having difficulties conceiving. Although our data did not include a sufficient number of cases to model primary infertility

separately, we did re-estimate the correlations, random effects models, and test-retest excluding primary infertility (that is, excluding non-pregnant women at parity 0). With one exception (discussed below), results were largely unchanged when women with primary infertility were excluded from the models.

For example, in the random effects models predicting unadjusted clinical infertility as a function of background sociodemographic characteristics (corresponding to Model 1 in Table 2 for the full sample), age (OR=1.09) and ethnicity (Denkyira OR=1.47) are the only significant predictors. The coefficient for age is identical to that in the full sample, while the odds ratio for Denkyira has increased by .02 (from OR=1.45 in the full sample). Similar patterns were found across the other measures and models—that is, occasional marginal shifts in the magnitude of coefficients, but no changes in significance levels or the direction of the relationships. The only exception was for the random effects models predicting unadjusted perceived difficulties: ethnicity became a marginally significant predictor in the sample restricted to women of parity greater than 0. Specifically, while the odds ratio for Adangbe was non-significant at 0.79 in the full sample, a stronger, significant association was observed for this ethnic group in the restricted sample (OR=0.64; p<0.05). The remaining coefficients in the restricted sample remained similar in significance, magnitude, and direction to those observed in the full sample.

Second, although we reported only results from the clinical, demographic 7 year, and self-identified measures for concision, it is possible that substantial differences exist between the two biomedical measures, as well as between the two demographic measures. To examine this possibility, we estimated both the test-retest and random effects models using the epidemiological and demographic-style 5 year measures (rather than the clinical and 7 year measures, as reported above). Results for the epidemiological models were nearly identical to those for the clinical measure, while results for the 5 year demographic measure were also very similar to those for the 7 year measure.

Discussion

A large array of infertility measures have been identified in the demographic and biomedical literature, contributing to lack of comparability across studies and widely varying prevalence rates (Gurunath et al., 2011). While we agree with the sentiment of previous calls for standardization of measurement (Gnoth et al., 2005; Gurunath et al., 2011; Habbema et al., 2004; Larsen, 2005), many of the proposed measures are of limited utility outside of a clinical setting. We contend that a suitable measure should be easily constructed from secondary survey data and should correlate with subjective assessments. In this paper, we used longitudinal survey data to answer the following: (1) which measure of infertility that can be constructed from survey data is tied to both objective biomedical/demographic measures and perceived difficulties conceiving and (2) which measures relate to background characteristics (age, ethnicity, religion) in an expected manner?

Capitalizing on the ability to assess within-person reliability of measures across time arising from the longitudinal data, results here show that while the trends in infertility across time are similar across the objective measures, longer waiting periods result in more stringent, stable measures. However, perceived difficulties conceiving were most closely aligned with clinical infertility, suggesting that long waiting times required by demographic-style measures may be too conservative to fit with women's own assessments. The demographic 7 year measure behaved in the most predictable way–that is, rising steadily across time, in keeping with bodily ageing and declining probability of conception. Notably, the fact that correlations between the same measure across waves are not higher likely reflects the idea that fertility is not a binary state, but a continuum along which conception may be easier or more difficult. Though there is a sporadic relationship between infertility, ethnicity, and religious affiliation, only age is consistently related to infertility. The models provide evidence that the effects of ethnicity and religious

affiliation are tied to one another, as well as being closely tied to contraceptive use, in keeping with literature outlining differences in sexual practices associated with ethnicity and religious beliefs (Frank, 1983).

In order to compare the same sample of women across measures while reconciling quite a disparate array of infertility criteria, we made several coding decisions about the measures which, in different instances, resulted in some modifications to the measures. First, we restricted the sample to women aged 15-49 because the relationship between self-identified difficulties conceiving and age (particularly cross-culturally) is not well-established in the literature, and because many survey data sets focussed on reproductive health tend to focus on this age range, likely reflecting both biological and normative limits on reproductive age around the globe. Second, we compared adjusted and unadjusted measures to examine the effect of fertility desires and contraceptive use. Gurunath et al. (2011) note that demographic-style measures of infertility often do not (or cannot) distinguish between voluntary and involuntary absence of a live birth, and so may overestimate infertility if the absence of a birth is voluntary. They suggest that marital status may be a useful proxy for voluntary childlessness depending on the context, but this suggestion of a context-specific measure seems to be at odds with the call for a standardized measure for cross-national comparisons across time. Furthermore, previous qualitative research has suggested that infertility may increase the risk of relationship disruption (Dyer, 2007; Dyer, Abrahams, Hoffman, & van der Spuy, 2002; Fledderjohann, 2012), pointing to potential endogeneity between marital status and infertility; empirical research is needed to test this association.

Here, we have examined how controlling for fertility desires and contraceptive use shapes our understanding of infertility. We found that the correlations revealed little difference between measures controlling for all contraceptive use compared to adjusted methods. Comparing unadjusted and adjusted measures reveals that a large proportion of those identified as infertile based purely on waiting times are

successful contraceptors; accounting for contraceptive use, then, is important to avoid false positives. Finally, we did not make any adjustments for postpartum amenorrhea, both because specific questions on amenorrhea may not be available in survey data, and because it may be difficult to distinguish postpartum from nutritional amenorrhea, which may be of particular concern in countries with high rates of malnutrition. As malnutrition may be sustained, nutritional amenorrhea could represent an barrier to fertility, and correction for this phenomenon could create a (likely slight) conservative bias.

In sum, demographic measures of infertility are too conservative to match with individual notions of difficulties conceiving. The distinction between controlling for modern versus all contraceptives made little difference; controlling for all methods is advisable for a stringent measure. Individual assessments of ability to conceive may be far more salient for understanding the social implications of infertility, and the clinical measure appears to be the most appropriate objective measure for social research on infertility. This likely reflects the short waiting time to conception often expected of couples. However, given the strong similarities between biomedical measures of infertility, epidemiological infertility may serve as a more reasonable and conservative substitute. The longer waiting time to conception required by an epidemiological measure is particularly important because, constructed on the basis of survey data using time to live births rather than conceptions, the clinical measure may identifies a woman as infertile after 12 months since her last birth. In reality, this leaves approximately 3 months for a conception to be achieved when 9 months of gestation is factored in. As a result, a clinical measure may overestimate infertility, especially as many women with potential secondary infertility will still be experiencing postpartum amenorrhea.

This possibility of overestimating infertility is particularly problematic for prevalence studies. While it would be possible with longitudinal survey data to retrospectively track the date of pregnancy based on the date of live birth and thereby in prove the measure, this method poses two problems for our

purposes. First, many of the most widely used data sources for this purpose (especially the Demographic and Health Surveys) are cross-sectional, and such retrospective correction would not be possible. Second, the method would be unlikely to capture spontaneous abortions, particularly those in early pregnancy, and inability to accurately track pregnancy wastage in survey data (J. B. Casterline, 1989) could result in a systematic bias. Furthermore, while a focus on ability to conceive is useful in a clinical setting, where treatment options vary based on the infertility etiology, arguably for most couples the ultimate goal is live birth, not a pregnancy. As infertility is a social and not just a biological condition, focus on a live birth is arguably preferable. As an epidemiological measure is less likely to overestimate infertility while still correlating closely with self-assessed measures, we recommend use of a 24 month epidemiological measure of infertility in future research, corrected for both fertility desires and contraceptive use.

There are some important limitations to the study. First, because the data are based solely on Ghana, the findings cannot be generalized more broadly. Additional research is needed both within and outside of SSA to determine if the measure recommended here performs well in other settings. Moreover, the sampling frame for the Cape Coast data was designed to maximize ethnic, economic, kinship, and between-community diversity (Casterline, 2007). This emphasis on diversity of the sample may have resulted in selection bias. Future research should attempt to replicate the findings using a nationally representative sample. Second, due to data limitations, it was not possible to distinguish between primary and secondary infertility. The prevalence of secondary infertility tends to be much higher than that of primary infertility (Larsen & Raggers, 2001; Mascarenhas et al., 2013), and the causes and social consequences of primary versus secondary infertility can be quite different (Larsen, 2000; R. G. White et al., 2001). Following from these points, it may also be the case that the reliability of infertility measurements, and particularly their relationship to background characteristics, may differ for primary and secondary infertility.

Third, in order to maximize criterion agreement between the objective measures of infertility used here, the versions of these measures in this paper are modified versions of the clinical, epidemiological, and demographic measures we identified in the literature. For instance, Larsen and Menken (Larsen & Menken, 1989, 1991) suggest using age 45 rather than age 49 as the upper bound age group for infertility calculations. In a similar vein, although a strict definition of biomedical infertility focuses on conception rather than live births, our analysis focused on births due to the difficulties measuring conception; a focus on conception may produce different results. However, it is also worth noting that all extant measures are imperfect indicators of underlying construct of ability to conceive, as evidenced, for example, by the by the large number of diagnosed cases of unexplained infertility. It is estimated that between 15 and 30% of clinical infertility cases are of unexplained etiology, and approximately 50% of these conceive within 12 months of this diagnosis (Gelbaya, Potdar, Jeve, & Nardo, 2014), suggesting that even clinical diagnosis is an imperfect indicator of ability to conceive.

Finally, it is unclear how the subjective measure of perceived difficulties conceiving may compare to other potential subjective measures. The U.S. National Survey of Fertility Barriers includes two questions which assess self-identified infertility directly: "Do you think you have/have had/might have trouble getting pregnant" and "Do you think you have/have had a fertility problem" (Johnson & White, n.d.). However, there is a serious dearth of survey data which includes subjective measures of infertility in any form. Much of the available data on subjective assessments of fertility status are qualitative, and consequently rely on small samples which are generally not publicly available. Future research should consider a variety of measures of perceived difficulties conceiving and self-identified infertility to assess which is most valid. In order to accomplish this, it is imperative that subjective measures of infertility be included in social surveys in SSA as a complement to existing qualitative work.

In sum, there is considerable heterogeneity in how infertility is measured and defined both within and between fields, which has been further complicated in SSA by the exclusion of infertility measures from survey data. There is no consensus on how best to measure infertility (Gnoth et al., 2005; Gurunath et al., 2011; Habbema et al., 2004). Infertility is a social problem as well as biological issue, and a subjective assessment is an important but previously neglected component of its measurement. As a result, while extant demographic measures offer an elegant and well-validated means of measuring prevalence based on cross-sectional survey data, they do not necessarily represent the best possible measure for other kinds of social research. Given the dearth of direct infertility measures in most data in SSA, we have sought here to identify a measure which closely matches standard objective measures, can be constructed from survey data, and, most importantly for our purpose, is correlated with respondents' own subjective assessments of their ability to conceive. Ultimately, our aim was not to definitively identify the best possible measure of infertility. Rather, we seek to identify the best measure given current data constraints and, importantly, to suggest that the social components of infertility need to be considered in its measurement. Additionally, we strongly encourage the widespread inclusion of subjective measures in survey data, and suggest that there is an urgent need for further development of subjective measures. Much work remains to be done on this important topic, but this work will be aided considerably by the inclusion of infertility measures in survey data.

References

Addai, I. (1999). Ethnicity and sexual behavior in Ghana. *Biodemography and Social Biology*, *46*(1-2), 17–32.

Addai, I. (2000). Religious affiliation and sexual initiation among Ghanaian women. *Review of Religious Research*, 328–343.

Addai, I., & Trovato, F. (1999). Structural assimilation and ethnic fertility in Ghana. *Journal of Comparative Family Studies*, *30*(3), 409–427.

Alwin, D. F. (2007). *Margins of error: a study of reliability in survey measurement* (Vol. 547). Hoboken, NJ: John Wiley & Sons, Inc.

Anarfi, J. K., & Owusu, A. Y. (2011). The making of a sexual being in Ghana: The state, religion and the influence of society as agents of sexual socialization. *Sexuality & Culture*, *15*(1), 1–18.

Andrews, F. M., Abbey, A., & Halman, L. J. (1991). Stress from infertility, marriage factors, and subjective well-being of wives and husbands. *Journal of Health and Social Behavior*. Retrieved from http://psycnet.apa.org/psycinfo/1992-24008-001

Armstrong, A. (1997). Law and the family in Southern Africa. In A. Adepoju (Ed.), *Family, population and development in Africa* (pp. 183–202). London: Zed Books.

Aryee, A. F. (1997). The African family and changing nuptiality patterns. In A. Adepoju (Ed.), *Family, population and development in Africa* (pp. 78–96). London: Zed Books.

Barden-O'Fallon, J. (2005). Unmet fertility expectations and the perception of fertility problems in a Malawian village. *African Journal of Reproductive Health*, 14–25.

Bledsoe, C. H. (2002). *Contingent lives: fertility, time, and aging in West Africa*. Chicago: University of Chicago Press.

Bongaarts, J., Frank, O., & Lesthaeghe, R. (1984). The proximate determinants of fertility in sub-Saharan Africa. *Population and Development Review*, *10*(3), 511–537.

Broekmans, F. J., Knauff, E. A., te Velde, E. R., Macklon, N. S., & Fauser, B. C. (2007). Female reproductive ageing: current knowledge and future trends. *Trends in Endocrinology & Metabolism*, *18*(2), 58–65.

Casterline, J. (2007). *Social learning, social influence, and fertility control* (Project description). New York: Population Council and University of Cape Coast.

Casterline, J. B. (1989). Collecting data on pregnancy loss: a review of evidence from the World Fertility Survey. *Studies in Family Planning*, *20*(2), 81–95.

Demographic and Health Surveys. (2011, February 25). Statistics compiler. Retrieved from http://www.statcompiler.com/country.cfm? ctryid=14&Ctryname=Ghana#

DeRose, L. F., & Ezeh, A. C. (2005). Men's influence on the onset and progress of fertility decline in Ghana, 1988–98. *Population Studies*, *59*(2), 197–210.

Dick, M.-L., Bain, C. J., Purdie, D. M., Siskind, V., Molloy, D., & Green, A. C. (2003). Self-reported difficulty in conceiving as a measure of infertility. *Human Reproduction*, *18*(12), 2711–2717.

Dodoo, F. N.-A., & Frost, A. E. (2008). Gender in African Population Research: The Fertility/Reproductive Health Example. *Annual Review of Sociology*, *34*(1), 431–452. http://doi.org/10.1146/annurev.soc.34.040507.134552

Donkor, E. S. (2008). Socio-cultural perceptions of infertility in Ghana. *Africa Journal of Nursing and Midwifery*, *10*(1), 22–34.

Dyer, S. J. (2007). The value of children in African countries - Insights from studies on infertility. *Journal of Psychosomatic Obstetrics and Gynecology*, *28*(2), 69–77.

Dyer, S. J., Abrahams, N., Hoffman, M., & van der Spuy, Z. M. van der. (2002). Men leave me as I cannot have children': women's experiences with involuntary childlessness. *Human Reproduction*, *17*(6), 1663–1668.

Dyer, S. J., Abrahams, N., Hoffman, M., & Van der Spuy, Z. M. (2002). Infertility in South Africa: women's reproductive health knowledge and treatment-seeking behaviour for involuntary childlessness. *Human Reproduction*, *17*(6), 1657–1662.

Dyer, S. J., Abrahams, N., Mokoena, N. E., Lombard, C. J., & van der Spuy, Z. M. (2005). Psychological distress among women suffering from couple infertility in South Africa: a quantitative assessment. *Human Reproduction*, *20*(7), 1938–1943.

Ericksen, K., & Brunette, T. (1996). Patterns and predictors of infertility among African women: a cross-national survey of twenty-seven nations. *Social Science & Medicine*, *42*(2), 209–220.

Ezeh, A. C. (1993). The influence of spouses over each other's contraceptive attitudes in Ghana. *Studies in Family Planning*, *24*(3), 163–174.

Farnes, C., Beckstrand, R. L., & Callister, L. C. (2011). Help-seeking behaviours in childbearing women in Ghana, West Africa. *International Nursing Review*, *58*(4), 491–497.

Fledderjohann, J. J. (2012). 'Zero is not good for me': implications of infertility in Ghana. *Human Reproduction*, *27*(5), 1383–1390. http://doi.org/10.1093/humrep/des035

Frank, O. (1983). Infertility in sub-Saharan Africa: estimates and implications. *Population and Development Review*, *9*(1), 137–144.

Gelbaya, T. A., Potdar, N., Jeve, Y. B., & Nardo, L. G. (2014). Definition and Epidemiology of Unexplained Infertility. *Obstetrical & Gynecological Survey*, *69*(2), 109–115.

Gerrits, T. (1997). Social and cultural aspects of infertility in Mozambique. *Patient Education and Counseling*, *31*(1), 39–48.

Gijsels, M., Mgalla, Z., & Wambura, L. (2001). 'No child to send': context and consequences of female infertility in northwest Tanzania. In J. T. Boerma & Z. Mgalla (Eds.), *Women and infertility in sub-Saharan Africa: A multidisciplinary perspective* (pp. 203–222). Amsterdam, NL: KIT Publishers.

Gnoth, C., Godehardt, E., Frank-Herrmann, P., Friol, K., Tigges, J., & Freundl, G. (2005). Definition and prevalence of subfertility and infertility. *Human Reproduction*, *20*(5), 1144–1147.

Greil, A. L. (1991). *Not yet pregnant: Infertile couples in contemporary America*. New Jersey: Rutgers University Press.

Greil, A. L., & McQuillan, J. (2004). Help-seeking patterns among subfecund women. *Journal of Reproductive and Infant Psychology*, *22*(4), 305–319.

Greil, A. L., McQuillan, J., Johnson, K., Slauson-Blevins, K., & Shreffler, K. M. (2010). The hidden infertile: infertile women without pregnancy intent in the United States. *Fertility and Sterility*, *93*(6), 2080–2083. http://doi.org/10.1016/j.fertnstert.2009.08.024

Gurunath, S., Pandian, Z., Anderson, R. A., & Bhattacharya, S. (2011). Defining infertility—a systematic review of prevalence studies. *Human Reproduction Update*, *17*(5), 575–588.

Gyimah, S. O. (2003). A cohort analysis of the timing of first birth and fertility in Ghana. *Population Research and Policy Review*, *22*(3), 251–266.

Habbema, J. D. F., Collins, J., Leridon, H., Evers, J. L., Lunenfeld, B., & others. (2004). Towards less confusing terminology in reproductive medicine: a proposal. *Human Reproduction*, *19*(7), 1497–1501.

Hall, D. B. (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*, *56*(4), 1030–1039.

Hartzel, J., Agresti, A., & Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling*, *1*(2), 81–102.

Hollos, M., & Larsen, U. (2008). Motherhood in sub-Saharan Africa: The social consequences of infertility in an urban population in northern Tanzania. *Culture, Health & Sexuality*, *10*(2), 159–173.

Hollos, M., Larsen, U., Obono, O., & Whitehouse, B. (2009). The problem of infertility in high fertility populations: Meanings, consequences and coping mechanisms in two Nigerian communities. *Social Science & Medicine*, *68*(11), 2061–2068. http://doi.org/10.1016/j.socscimed.2009.03.008

Jacob, M. C., McQuillan, J., & Greil, A. L. (2007). Psychological distress by type of fertility barrier. *Human Reproduction*, *22*(3), 885–894.

Johnson, D. R. (1995). Assessing marital quality in longitudinal and life course studies. In J. Conoley & E. Buterick Werth (Eds.), *Family assessment* (pp. 155–202). Lincoln, NE: Buros Institute of Mental Measurements. Retrieved from http://digitalcommons.unl.edu/burosfamily/10/

Johnson, D. R., & White, L. (n.d.). *National Survey of Fertilty Barriers [Computer File]*. University Park, PA: The Pennsylvania State University Population Research Institute [distributor].

Koster-Oyekan, W. (1999). Infertility among Yoruba women: Perceptions on causes, treatments and consequences. *African Journal of Reproductive Health*, *3*(1), 13–26.

Larsen, U. (1994). Sterility in sub-Saharan Africa. *Population Studies*, *48*(3), 459–474.

Larsen, U. (2000). Primary and secondary infertility in sub-Saharan Africa. *International Journal of Epidemiology*, *29*(2), 285–291.

Larsen, U. (2005). Research on infertility: Which definition should we use? *Fertility and Sterility*, *83*(4), 846–852. http://doi.org/10.1016/j.fertnstert.2004.11.033

Larsen, U., & Menken, J. (1989). Measuring sterility from incomplete birth histories. *Demography*, *26*(2), 185–201.

Larsen, U., & Menken, J. (1991). Individual-level sterility: a new method of estimation with application to sub-Saharan Africa. *Demography*, *28*(2), 229–247.

Larsen, U., & Raggers, H. (2001). Levels and trends in infertility in sub-Saharan Africa. In J. T. Boerma & Z. Mgalla (Eds.), *Women and infertility in sub-Saharan Africa: A multidisciplinary perspective* (pp. 25–70). Amsterdam, NL: KIT Publishers.

Leonard, L. (2002). Problematizing fertility: 'Scientific' accounts and Chadian women's narratives. In M. Inhorn & F. Van Balen (Eds.), *Infertility around the globe: new thinking on childlessness, gender, and reproductive technologies* (pp. 193–214). Berkeley: University of California Press.

Leridon, H. (1991). [Sterility and sub-fecundity: from silence to impatience?]. *Population*, *46*(2), 225–47.

Loftus, J. (2009). 'Oh, No, I'm Not Infertile': Culture, Support Groups, and the Infertile Identity. *Sociological Focus*, *42*(4), 394–416.

Marchbanks, P. A., Peterson, H. B., Rubin, G. L., & Wingo, P. A. (1989). Research on infertility: definition makes a difference. *American Journal of Epidemiology*, *130*(2), 259–267.

Mascarenhas, M. N., Flaxman, S. R., Boerma, T., Vanderpoel, S., Mathers, C. D., & Stevens, G. A. (2013). Trends in primary and secondary infertility prevalence since 1990: a systematic analysis of demographic and reproductive health surveys. *The Lancet*, *381*, S90.

Mayaud, P. (2001). The role of reproductive tract infections. In J. T. Boerma & Z. Mgalla (Eds.), *Women and infertility in sub-Saharan Africa: A multidisciplinary perspective* (pp. 71–108). Amsterdam, NL: KIT Publishers.

Meekers, D., & Calvès, A.-E. (1997). 'Main'girlfriends, girlfriends, marriage, and money: The social context of HIV risk behaviour in sub-Saharan Africa. *Health Transition Review*, *7 (supplement)*, 361–375.

Mgalla, Z., & Boerma, J. T. (2001). The discourse of infertility in Tanzania. In J. T. Boerma & Z. Mgalla (Eds.), *Women and infertility in sub-Saharan Africa: A multidisciplinary perspective* (pp. 189–202). Amsterdam, NL: KIT Publishers.

Oppong, C., & Abu, K. (1987). *Seven roles of women: impact of education, migration and employment on Ghanaian mothers*. Geneva: International Labour Office.

Population Reference Bureau. (2012). *World population data sheet*. Population Reference Bureau. Retrieved from http://www.prb.org/Publications/Datasheets/2012/world-population-data-sheet/data-sheet.aspx

Rouchou, B. (2013). Consequences of infertility in developing countries. *Perspectives in Public Health*, *133*(3), 174–179. http://doi.org/10.1177/1757913912472415

Tabong, P. T.-N., & Adongo, P. B. (2013). Understanding the social meaning of infertility and childbearing: a qualitative study of the perception of childbearing and childlessness in Northern Ghana. *PloS One*, *8*(1), e54429.

Takyi, B. K., & Gyimah, S. O. (2007). Matrilineal family ties and marital dissolution in Ghana. *Journal of Family Issues*, *28*(5), 682–705.

Weinstein, M., Wood, J., & Chang, M. (1993). Age patterns of fecundability. Retrieved from http://www.popline.org/node/324690

White, M., Muhidin, S., Andrzejewski, C., Tagoe, E., Knight, R., & Reed, H. (2008). Urbanization and fertility: An event-history analysis of coastal Ghana. *Demography*, *45*(4), 803–816.

White, R. G., Zaba, B., Boerma, J. T., & Blacker, J. (2001). Modeling the dramatic decline of primary infertility in sub-Saharan Africa. In J. T. Boerma & Z. Mgalla (Eds.), *Women and infertility in sub-Saharan Africa: A multidisciplinary perspective* (pp. 117–150). Amsterdam, NL: KIT Publishers.

Table 1. Proportion Infertile Across Waves

Variable Name	W1	W2	W3	W4	W5	W6	W7	W8
Unadjusted Clinical	0.63	0.67	0.67	0.65	0.73	0.74	0.73	0.69
Adjusted Clinical	0.16	0.12	0.13	0.14	0.16	0.17	0.18	0.17
Unadjusted Demographic 7 Year	0.19	0.19	0.20	0.19	0.20	0.19	0.21	0.24
Adjusted Demographic 7 Year	0.03	0.01	0.01	0.02	0.02	0.02	0.03	0.04
Unadjusted Perceived Difficulties	0.66	--	--	--	--	0.66	0.60	0.65
Adjusted Perceived Difficulties	0.20	--	--	--	--	0.18	0.15	0.20

N=1,350

Table 2. Random effects model of clinical infertility, demographic infertility, perceived difficulties conceiving, and background characteristics, Cape Coast data, Ghana, 1998-2004

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	Unadjusted Clinical		Adjusted Clinical		Unadjusted Demographic		Adjusted Demographic		Unadjusted Perceived Difficulties		Adjusted Perceived Difficulties	
	OR	St. Error	OR	St. Error	OR	St. Error	OR	St. Error	OR	St. Error	OR	St. Error
Age	1.09***	0	1.17***	0.05	1.18***	0.01	1.40**	0.11	0.74***	0.02	0.91***	0.01
Age squared	--	--	0.02***	0.02	--	--	0.01**	0.01	88.14***	40.2	--	--
Ethnicity	--	--	--	--	--	--	--	--	--	--	--	--
Adangbe	0.96	0.13	1.21	0.2	1.2	0.25	2.16	1.03	0.79	0.12	0.93	0.23
Ga or Ewe	1.2	0.14	1.04	0.18	1.32	0.23	1.17	0.43	0.88	0.13	0.97	0.18
Denkyira	1.45***	0.18	1.07	0.15	1.75*	0.37	1.29	0.58	1.16	0.18	1.03	0.2
Fante	1.06	0.11	0.92	0.11	1.28	0.2	0.91	0.33	0.99	0.11	0.86	0.13
Ahanta or Other (ref)	--	--	--	--	--	--	--	--	--	--	--	--
Religious Affiliation	--	--	--	--	--	--	--	--	--	--	--	--
Catholic	1.07	0.17	0.97	0.18	1.25	0.22	0.59	0.25	0.74*	0.11	1.1	0.29
Protestant	1.05	0.15	0.95	0.18	1.52**	0.23	0.9	0.28	0.99	0.14	1.17	0.3
Moslem	1.03	0.18	0.82	0.16	1.38	0.33	0.53	0.24	0.8	0.14	0.97	0.28
Pentecostal or Charismatic	0.98	0.16	0.98	0.15	1.26	0.23	0.63	0.21	0.88	0.12	1.16	0.26
Syncretic, Traditional, or Other	0.89	0.12	0.89	0.14	1.01	0.19	0.46	0.18	0.96	0.15	1.1	0.25
None (ref)	--	--	--	--	--	--	--	--	--	--	--	--

Pooled N=10,800; *p<.05; **p<.01; ***p<.001; Age squared divided by 1,000

Table 3. Test-retest model for clinical, demographic 7 year, and perceived difficulties conceiving, Cape Coast data, Ghana, 1998-2004

	B	St. Error
Latent infertility W6	–	–
Clinical	1.01	0.03
Demographic	0.88***	0.03
Perceived difficulties conceiving	0.86***	0.03
Latent infertility W7	–	–
Clinical	1.01	0.04
Demographic	0.88***	0.04
Perceived difficulties conceiving	0.86***	0.04
Latent infertility W8	–	–
Clinical	1.02	0.03
Demographic	0.90***	0.03
Perceived difficulties conceiving	0.88***	0.04
Latent infertility W6 on W7	0.57***	0.06
Latent infertility W7 on W8	0.58***	0.06
Correlation of errors of clinical W6 and W7	0.01	0.06
Correlation of errors of clinical W6 and W8	0.05	0.06
Correlation of errors of clinical W7 and W8	-0.06	0.06
Correlation of errors of demographic W6 and W7	0.32***	0.08
Correlation of errors of demographic W6 and W8	0.41***	0.11
Correlation of errors of demographic W7 and W8	0.36***	0.06
Correlation of errors of perceived diff. W6 and W7	0.1	0.07
Correlation of errors of perceived diff. W6 and W8	0.19***	0.06
Correlation of errors of perceived diff. W7 and W8	0.06	0.07

Notes: N=1,350; *p < .05; **p <.01; ***p < .001; RMSEA=.027

Supplemental Materials

Supplemental Figure 1. Data collection schedule and sample sizes by wave for the Cape Coast data, 1998-2004

Wave 1, Oct. 1998-Feb. 1999
[n=1,219]

↓

Wave 2, May-Aug. 1999
[n=1,010 (+209 supplement)=1,364]

↓

Wave 3, Feb.-May 2000
[n=1,310]

↓

Wave 4, Sept.-Nov. 2000
[n=1,312]

↓

Wave 5, Feb.-May 2001
[n=1,284]

↓

Wave 6, Sept.-Dec. 2001
[n=1,253]

↓

Wave 7, July-Nov. 2002
[n=1,241]

↓

Wave 8, Nov. 2003-Feb. 2004
[n=1,218]

Supplemental Table 1. Infertility measures by waiting time and contraceptive use

Measure Name	Minimum Waiting Time	Contraceptive Users Coded As Not Infertile
Unadjusted Clinical	12 months	None
Contraceptive Adjusted Clinical	12 months	All
Modern Adjusted Clinical	12 months	Modern Users Only
Unadjusted Epidemiological	24 Months	None
Contraceptive Adjusted Epidemiological	24 Months	All
Modern Adjusted Epidemiological	24 Months	Modern Users Only
Unadjusted Demographic 5 Year	60 Months	None
Contraceptive Adjusted Demographic 5 Year	60 Months	All
Modern Adjusted Demographic 5 Year	60 Months	Modern Users Only
Unadjusted Demographic 7 Year	84 Months	None
Contraceptive Adjusted Demographic 7 Year	84 Months	All
Modern Adjusted Demographic 7 Year	84 Months	Modern Users Only
Unadjusted Perceived Difficulties	None	None
Contraceptive Adjusted Perceived Difficulties	None	All
Modern Adjusted Perceived Difficulties	None	Modern Users Only

Supplemental Table 2. Descriptive Statistics

Variable Name	Mean	St. Dev	Min	Max
Age	32.33	8.63	15.00	49.00
Age at First Marriage	19.47	3.33	10.00	31.00
Ethnicity	--	--	--	--
Adangbe	0.16	0.37	0.00	1.00
Ga or Ewe	0.11	0.31	0.00	1.00
Denkyira	0.13	0.33	0.00	1.00
Fante	0.51	0.50	0.00	1.00
Ahanta or Other	0.10	0.30	0.00	1.00
Marital Status	--	--	--	--
Married	0.62	0.48	0.00	1.00
Single	0.12	0.33	0.00	1.00
In a Union	0.15	0.35	0.00	1.00
Separated	0.03	0.18	0.00	1.00
Divorced	0.05	0.22	0.00	1.00
Widowed	0.02	0.15	0.00	1.00
Education Level	--	--	--	--
No Education	0.36	0.48	0.00	1.00
Some Primary School	0.18	0.39	0.00	1.00
Finished Primary School	0.06	0.24	0.00	1.00
Attended Middle School	0.36	0.48	0.00	1.00
Attended Secondary School	0.04	0.19	0.00	1.00
Scale of Household Goods	2.95	2.13	0.00	9.00
Religious Affiliation	--	--	--	--
Catholic	0.15	0.36	0.00	1.00
Protestant	0.19	0.39	0.00	1.00
Moslem	0.22	0.42	0.00	1.00
Pentecostal or Charismatic	0.13	0.34	0.00	1.00
Syncretic, Traditional, or Other	0.22	0.41	0.00	1.00
None	0.08	0.27	0.00	1.00
Desire Additional Children	0.57	0.50	0.00	1.00
Using Any Birth Control	0.65	0.48	0.00	1.00
Using Modern Birth Control	0.57	0.50	0.00	1.00
Parity	3.53	2.90	0.00	14.00
Time to Pregnancy	--	--	--	--
Quick	0.36	0.48	0.00	1.00
Takes a Long Time	0.32	0.47	0.00	1.00
Impossible	0.06	0.24	0.00	1.00
Don't Know	0.26	0.44	0.00	1.00

N=1,350; Pooled N=10,800

Supplemental Table 3. Proportion Infertile Across Waves

Variable Name	W1	W2	W3	W4	W5	W6	W7	W8
Unadjusted Clinical	0.63	0.67	0.67	0.65	0.73	0.74	0.73	0.69
Adjusted Clinical	0.16	0.12	0.13	0.14	0.16	0.17	0.18	0.17
Unadjusted Epidemiological	0.52	0.52	0.52	0.47	0.54	0.54	0.60	0.64
Adjusted Epidemiological	0.12	0.08	0.08	0.08	0.11	0.12	0.13	0.15
Unadjusted Demographic 5 Year	0.27	0.27	0.27	0.26	0.29	0.27	0.29	0.35
Adjusted Demographic 5 Year	0.05	0.02	0.03	0.03	0.04	0.04	0.04	0.07
Unadjusted Demographic 7 Year	0.19	0.19	0.20	0.19	0.20	0.19	0.21	0.24
Adjusted Demographic 7 Year	0.03	0.01	0.01	0.02	0.02	0.02	0.03	0.04
Unadjusted Perceived Difficulties	0.66	--	--	--	--	0.66	0.60	0.65
Adjusted Perceived Difficulties	0.20	--	--	--	--	0.18	0.15	0.20

N=1,350